Splitting for Rare Event Simulation: A Large Deviations Approach to Design and Analysis

Thomas Dean* and Paul Dupuis†
Lefschetz Center for Dynamical Systems
Division of Applied Mathematics
Brown University

Abstract

Particle splitting methods are considered for the estimation of rare events. The probability of interest is that a Markov process first enters a set B before another set A, and it is assumed that this probability satisfies a large deviation scaling. A notion of subsolution is defined for the related calculus of variations problem, and two main results are proved under mild conditions. The first is that the number of particles generated by the algorithm grows subexponentially if and only if a certain scalar multiple of the importance function is a subsolution. The second is that, under the same condition, the variance of the algorithm is characterized (asymptotically) in terms of the subsolution. The design of asymptotically optimal schemes is discussed, and numerical examples are presented.

1 Introduction

The numerical estimation of probabilities of rare events is a difficult problem. There are many potential applications in operations research and engineering, insurance, finance, chemistry, biology, and elsewhere, and many papers (and by now even a few books) have proposed numerical schemes for particular settings and applications. Because the quantity of interest is very small, standard Monte Carlo simulation requires an enormous number of samples

*Research of this author supported in part by the National Science Foundation (NSF-DMS-0306070 and NSF-DMS-0404806).

†Research of this author supported in part by the National Science Foundation (NSF-DMS-0306070 and NSF-DMS-0404806) and the Army Research Office (W911NF-05-1-0289).

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2007		**00-00-2007 to 00-00-2007**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Splitting for Rare Event Simulation: A Large Deviations Approach to Design and Analysis	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Brown University,Lefschetz Center for Dynamical Systems,182 George Street,Providence,RI,02912**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Particle splitting methods are considered for the estimation of rare events. The probability of interest is that a Markov process first enters a set B before another set A, and it is assumed that this probability satisfies a large deviation scaling. A notion of subsolution is defined for the related calculus of variations problem, and two main results are proved under mild conditions. The first is that the number of particles generated by the algorithm grows subexponentially if and only if a certain scalar multiple of the importance function is a subsolution. The second is that, under the same condition, the variance of the algorithm is characterized (asymptotically) in terms of the subsolution. The design of asymptotically optimal schemes is discussed, and numerical examples are presented.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**34**	

for the variance of the resulting estimate to be comparable to the unknown probability. It quickly becomes unusable, and more efficient alternatives are sought.

The two most widely considered alternatives are those based on change-of-measure techniques and those based on branching processes. The former is usually called *importance sampling*, and the latter is often referred to as *multi-level splitting*. While good results on a variety of problem formulations have been reported for both methods, it is also true that both methods can produce inaccurate and misleading results. The design issue is critical, and one can argue that the proper theoretical tools for the design of importance sampling and splitting algorithms were simply not available for complicated models and problem formulations.

Suppose that the probability of interest takes the form $p = P\{Z \in G\} = \mu(G)$, where G is a subset of some reasonably regular space (e.g., a Polish space S) and μ a probability measure. In ordinary Monte Carlo one generates a number of independent and identically distributed (iid) samples $\{Z_i\}$ from μ, and then estimates p using the sample mean of $1_{\{Z_i \in G\}}$. In the case of importance sampling, one uses an alternative sampling distribution ν, generates iid samples $\left\{\bar{Z}_i\right\}$ from ν, and then estimates via the sample mean of $[d\mu/d\nu]\,(\bar{Z}_i)1_{\left\{\bar{Z}_i \in G\right\}}$. The Radon-Nikodim derivative $[d\mu/d\nu]\,(\bar{Z}_i)$ guarantees that the estimate is unbiased. The goal is to choose ν so that the individual samples $[d\mu/d\nu]\,(\bar{Z}_i)1_{\left\{\bar{Z}_i \in G\right\}}$ cluster tightly around p, thereby reducing the variance. However, for complicated process models or events G the selection of a good measure ν may not be simple. The papers [9, 10] show how certain standard heuristic methods based on ideas from large deviations could produce very poor results. The difficulty is due to points in S with low probability under ν for which $d\mu/d\nu$ is very large. The aforementioned large deviation heuristic does not properly account for the contribution of these points to the variance of the estimate, and it is not hard to find examples where the corresponding importance sampling estimator is much worse than even ordinary Monte Carlo. The estimates exhibit very inaccurate and/or unstable behavior, though the instability may not be evident from numerical data until massive amounts have been generated.

The most discussed application of splitting type schemes is to first entrance probabilities, and to continue the discussion we specialize to that case. Thus Z is the sample path of a stationary stochastic process $\{X_i\}$ (which for simplicity is taken to be Markovian), and G is the set of trajectories that first enter a set B prior to entering a set A. More precisely, for

disjoint B and A and $x \notin A \cup B$,

$$p = p(x) = P\left\{X_j \in B, X_i \notin A, i \in \{0, \ldots, j\}, j < \infty | X_0 = x\right\}.$$

In the most simple version of splitting, the state space is partitioned according to certain sets $B \subset C_0 \subset C_1 \subset \cdots \subset C_K$, with $x \notin C_K$ and $A \cap C_K = \emptyset$. These sets are often defined as level sets of a particular function V, which is commonly called an *importance function*. Particles are generated and killed off according to the following rules. A single particle is started at x. Generation of particles (splitting) occurs whenever an existing particle reaches a threshold or level C_i for the first time. At that time, a (possibly random) number of new particles are placed at the location of entrance into C_i. The future evolutions of these particles are independent of each other (and all other particles), and follow the law of $\{X_i\}$. Particles are killed if they enter A before B. Attached to each particle is a weight. Whenever a particle splits the weight of each descendent equals that of the parent times a discount factor. A random tree is thereby produced, with each leaf corresponding to a particle that has either reached B or been killed. A random variable (roughly analogous to a single sample $[d\mu/d\nu]\,(\bar{Z}_i)1_{\left\{\bar{Z}_i \in G\right\}}$ from the importance sampling approach) is defined as the sum of the weights for all particles that make it to B. The rule that updates the weights when a particle splits is chosen so that the expected value of this random variable is p. This numerical experiment is independently repeated a number of times, and the sample mean is again used to estimate p.

There are two potential sources of poor behavior in the splitting algorithm. The first and most troubling is that the number of particles may be large. For example, the number could be comparable θ^K for some $\theta > 1$. In settings where a large deviation characterization of p is available, the number of levels itself usually grows with the large deviation parameter, and so the number of particles could grow exponentially. We will refer to this as *instability* of the algorithm. For obvious computational reasons, instability is something to be avoided. The other source of poor behavior is analogous to that of importance sampling (and ordinary Monte Carlo), which is high relative variance of the estimate. If the weighting rule leads to high variation of the weights of particles that make it to B, or if too many simulations produce no particles that make it to B (in which case a zero is averaged in the sample mean), then high relative variance is likely. Note, however, that this problem has a bounded potential for mischief, since the weights cannot be larger than one. Such a bound does not hold for the Radon-Nikodim derivative of importance sampling.

When the probability of interest can be approximated via large deviations, the rate of decay is described in terms of a variational problem, such as a calculus of variations or optimal control problem. It is well known that problems of this sort are closely related to a family of nonlinear partial differential equations (PDE) known as Hamilton-Jacobi-Bellman (HJB) equations. In a pair of recent papers [3, 5], it was shown how subsolutions of the HJB equations associated with a variety of rare event problems could be used to construct and rigorously analyze efficient importance sampling schemes. In fact, the subsolution property turns out to be in some sense necessary and sufficient, in that efficient schemes can be shown to imply the existence of an associated subsolution.

The purpose of the present paper is to show that in certain circumstances a remarkably similar result holds for splitting algorithms. More precisely, we will show the following under relatively mild conditions.

- A necessary and sufficient condition for the stability of the splitting scheme associated to a given importance function is that a certain scalar multiple of the importance function be a subsolution of the related HJB equation. The multiplier is the ratio of the logarithm of the expected number of offspring for each split and the gap between the levels.

- If the subsolution property is satisfied, then the variance of the splitting scheme decays exponentially with a rate defined in terms of the value of the subsolution at a certain point.

- As in the case of importance sampling, when a subsolution has the maximum possible value at this point (which is the value of the corresponding solution), the scheme is in some sense asymptotically optimal.

These results are significant for several reasons. The most obvious is that a splitting algorithm is probably not useful if it is not stable, and the subsolution property provides a way of checking stability. A second is that good, suboptimal schemes can be constructed and compared via the subsolutions framework. A third reason is that for interesting classes of problems it is possible to construct subsolutions that correspond to asymptotically optimal algorithms (see [3, 5]). Subsolutions can be much easier to construct than solutions. In this context it is worth noting that the type of subsolution required for splitting (a viscosity subsolution [1, 6]) is less restrictive that

the type of subsolution required for importance sampling. Further remarks on this point will be given in Section 5.

An outline of the paper is as follows. In the next section we describe the probabilities to be approximated, state assumptions, and formulate the splitting algorithm. This section also presents a closely related algorithm that will be used in the analysis. Section 3 studies the stability problem, and Section 4 shows how to bound the variance of an estimator in terms of the related subsolution. The results of Sections 3 and 4 can be phrased directly in terms of the solution to the calculus of variations problem that is related to the large deviation asymptotics. However, for the purposes of practical construction of importance functions the characterization via subsolutions of a PDE is more useful. These issues are discussed in Section 5, and examples and numerical examples are presented in the concluding Section 6.

Acknowledgment. Our interest in the parallels between importance sampling and multi-level splitting was stimulated by a talk given by P.T. de Boer at the RESIM conference in Bamberg, Germany [2].

2 Problem Formulation

2.1 Problem Setting and Large Deviation Properties

A domain $D \subset \mathbb{R}^d$ is given and also a sequence of discrete time, stationary, Markov $D-$valued processes $\{X^n\}$. Disjoint sets A and B are given, and we set $\tau^n \doteq \min\{i : X_i^n \in A \cup B\}$. The probability of interest is then

$$p^n(x_n) \doteq P\left\{X_{\tau^n}^n \in B \,|X_0^n = x_n\right\}.$$

The varying initial conditions are used for greater generality, but also because initial conditions for the prelimit processes may be restricted to some subset of D. The analogous continuous time framework can also be used with analogous assumptions and results. For a given point $x \notin A \cup B$, we make the following large deviation-type assumption.

Condition 1 *For any sequence $x_n \to x$,*

$$\lim_{n\to\infty} -\frac{1}{n}\log p^n(x_n) = W(x),$$

where $W(x)$ is the solution to a control problem of the form

$$\inf \int_0^t L\left(\phi(s), \dot{\phi}(s)\right) ds.$$

Here $L : \mathbb{R}^d \times \mathbb{R}^d \to [0,\infty]$, and the infimum is taken over all absolutely continuous functions ϕ with $\phi(0) = x$, $\phi(t) \in B, \phi(s) \notin A$ for all $s \in [0,t]$ and some $t < \infty$.

Remark 2 The assumption that $\{X^n\}$ be Markovian is not necessary for the proofs to follow. For example, it could be the case that X_i^n is the first component of a Markov process (X_i^n, Y_i^n) (e.g., so-called Markov-modulated processes). In such a case it is enough that the analogous large deviation limit hold uniformly in all possible initial conditions Y_0^n, and indeed the proofs given below will carry over with only notational changes. This can be further weakened, e.g., it is enough that the estimates hold uniformly with sufficiently high probability in the conditioning data. However, the construction of subsolutions will be more difficult, since the PDE discussed in Section 5 is no longer available in explicit form. See [5] for further discussion on this point.

It is useful to say a few words on how one can verify conditions like Condition 1 from existing large deviation results. Similar but slightly different assumptions will be made in various places in the sequel, and in all cases analogous remarks will apply.

For discrete time processes one often finds process-level large deviation properties phrased in terms of a continuous time interpolation $X^n(t)$, with $X^n(i/n) \doteq X_i^n$ and $X^n(t)$ defined by piecewise linear interpolation for t not of the form $t = i/n$. In precise terms, process-level large deviation asymptotics hold for $\{X^n\}$ if the following upper and lower bounds hold for each $T \in (0,\infty)$ and any sequence of initial conditions $x_n \in D$ with $x_n \to x$. Define

$$I_x^T(\phi) = \int_0^T L\left(\phi(s), \dot{\phi}(s)\right) ds$$

if ϕ is absolutely continuous with $\phi(0) = x$, and $I_x^T(\phi) = \infty$ otherwise. If F is any closed subset of $C([0,T] : D)$ then the upper bound

$$\limsup_{n\to\infty} \frac{1}{n} \log P\left\{X^n \in F \,|X^n(0) = x_n\right\} \leq -\inf_{\phi\in F} I_x^T(\phi)$$

holds, and if O is any open subset of $C([0,T] : D)$ then the lower bound

$$\liminf_{n\to\infty} \frac{1}{n} \log P\left\{X^n \in O \,|X^n(0) = x_n\right\} \geq -\inf_{\phi\in O} I_x^T(\phi)$$

holds. It is also usual to assume that for each fixed x, T, and any $M < \infty$, the set

$$\left\{\phi \in C([0,T] : D) : I_x^T(\phi) \le M\right\}$$

is compact in $C([0,T] : D)$. The zero-cost trajectories (i.e., paths ϕ for which $I_x^T(\phi) = 0$) are particularly significant in that all other paths are in some sense exponentially unlikely.

With regard to Condition 1, two different types of additional conditions beyond the sample path large deviation principle are required. One is a condition that allows a reduction to large deviation properties over a finite time interval. For example, suppose that there is $\bar{T}$ such that if ϕ enters neither A nor B before $\bar{T}$, then $I_x^{\bar{T}}(\phi) \ge W(x) + 1$. In this case, the contribution to $p^n(x_n)$ from sample paths that take longer than $\bar{T}$ is negligible, and can be ignored. This allows an application of the finite time large deviation principle. Now let G be the set of trajectories that enter B at some time $t < \bar{T}$ without having previously entered A. By the first condition, the asymptotic rates of decay of $p^n(x_n)$ and $P\left\{X^n \in G \,|X^n(0) = x_n\right\}$ are the same. The second type of condition is to impose enough regularity on the sets A and B and the rate function $I_x^T(\phi)$ that the infimum over the interior and closure of G are the same. These points are discussed at length in the literature on large deviations [7].

Example 3 *Assume the following conditions: $L(\cdot,\cdot)$ is lower semicontinuous; for each $x \in D, L(x,\cdot)$ is convex; $L(x,\cdot)$ is uniformly superlinear; for each $x \in D$ there is a unique point $b(x)$ for which $L(x, b(x)) = 0$; b is Lipschitz continuous, and all solutions to $\dot{\phi} = b(\phi)$ are attracted to $\theta \in A$, with A open. Let $\mathcal{D} \subset D$ be a bounded domain that contains A and B, and assume $\langle b(x), n(x)\rangle < 0$ for $x \in \partial\mathcal{D}$, where $n(x)$ is the outward normal to $\mathcal{D}$ at x. Suppose that the cost to go from x to any point in $\partial\mathcal{D}$ is at least $W(x) + 1$. Then $\bar{T}$ as described above exists.*

2.2 The Splitting Algorithm

In order to define a spitting algorithm we need to choose an importance function $V(y)$ and a level size $\Delta > 0$. We will require that $V(y)$ be continuous and that $V(y) \le 0$ for all $y \in B$. Later on we will relate V to the value function W, and discuss why subsolutions to the PDE that is satisfied by W are closely related to natural candidates for the importance function.

To simplify the presentation, we consider only splitting mechanisms with an a priori bound $R < \infty$ on the maximum number of offspring. The restriction is convenient for the analysis, and as we will see is without loss of

generality. The set of (deterministic) splitting mechanisms will be indexed by $j \in \{1, \ldots, J\}$. Given that mechanism j has been selected, $r(j)$ particles (with $|r(j)| \leq R$) are generated and weights $w(j) \in \mathbb{R}_+^{r(j)}$ are assigned to the particles. Note that we do not assume $\sum_{i=1}^{r(j)} w_i(j) = 1$. The class of all splitting mechanisms (i.e., including randomized mechanisms) is identified with the set of all probability distributions on $\{1, \ldots, J\}$.

Associated with V are the level sets

$$L_z = \{y \in D : V(y) \leq z\}.$$

A key technical condition we use is the following. In the condition, E_x denotes expected value given $X_0^n = x$.

Condition 4 *Let $z \in [0, V(x)]$ be given and define $\sigma^n \doteq \min\{i : X_i^n \in A \cup L_z\}$. Then*

$$\liminf_{n\to\infty} -\frac{1}{n} \log E_{x_n}\left[1_{\{X_{\sigma^n}^n \in L_z\}} \left(p^n(X_{\sigma^n}^n)\right)^2\right] \geq W(x) + \inf_{y\in\partial L_z} W(y).$$

Under the conditions discussed after Condition 1 which allow one to consider bounded time intervals, Condition 4 follows from the Markov property and the large deviation upper bound.

We also define collections of sets $\left\{C_0^n = B, C_j^n = L_{(j-1)\Delta/n}, j = 1, \ldots\right\}$. Define the level function l^n by $l^n(y) \doteq \min\{j \geq 0 : y \in C_j^n\}$. The location of the starting point corresponds to $l^n(x) = \lceil nV(x)/\Delta \rceil$, and $l^n = 0$ indicates entry into the target set B. The splitting algorithm associated with a particular distribution q will now be defined. Although the algorithm depends on $V, q, r, w, x_n, \Delta, A$ and B, to minimize notational clutter these dependencies are not explicitly denoted.

Splitting Algorithm (SA)

```
Variables:
   N_r^n number of particles in generation r
   X_{r,k}^n position of k^th particle in generation r
   w_{r,k}^n weight of k^th particle in generation r
Initialization Step:
   N_0^n = 1, X_{0,1}^n = x_n, w_{0,1}^n = 1
   for r = 1, ..., l^n(x_n)
      N_r^n = 0
   end
```

```
Main Algorithm:
   for r = 1, ..., l^n(x_n)
      if N^n_{r-1} = 0 then N^n_r = 0
      else
         for j = 1, ..., N^n_{r-1}
            generate Z^n_{r,j,i} a single sample of a process with the
            same law as X^n_i and initial condition Z^n_{r,j,0} = X^n_{r-1,j}

            let τ^n_{r,j} = inf{i : Z^n_{r,j,i} ∈ A ∪ C^n_{l^n(x_n)-r}}

            Splitting Step begin
            if Z^n_{r,j,τ^n_{r,j}} ∉ A
               let M be an independent sample from the law q
               for k = 1, ..., |r(M)|
                  N^n_r = N^n_r + 1
                  X^n_{r,N^n_r} = Z^n_{r,j,τ^n_{r,j}}
                  w^n_{r,N^n_r} = w_k(M) w^n_{r-1,j}
               end
            end
            Splitting Step end

         end
      end
   end
Construction of a sample:
   once all the generations have been calculated we form
   the quantity
      s^n_SA = Σ_{j=1}^{N^n_{l^n(x_n)}} w^n_{l^n(x_n),j} .
```

An estimate $\hat{p}^n_{\mathrm{SA}}(x_n)$ of $p^n(x_n)$ is formed by averaging a number of independent samples of s^n_{SA}. Observe that once generation r has been calculated the information about generations 0 to $r-1$ can be discarded, and so there is no need to keep all the data in memory until completion of the algorithm. Also note that in practice there is no need to split upon entering $C^n_0 = B$.

We first need to find conditions under which this splitting algorithm gives an unbiased estimator of $p^n(x_n)$. To simplify this and other calculations we

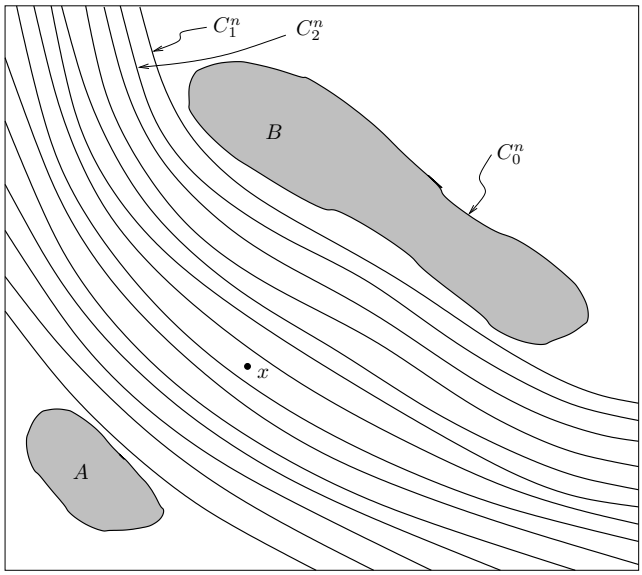

Figure 1: The Sets A and B and Level Sets of V.

introduce an auxiliary algorithm titled Splitting Algorithm Fully Branching (SFB). The essential difference between the two is that the process dynamics are redefined in A to make it absorbing, and that splitting continues even after a particle enters A. When the estimate is constructed we only count the particles which are in B in the last generation, so that the two estimates have the same distribution. The SFB algorithm is more convenient for purposes of analysis, because we do not distinguish those particles which have entered A from those which still have a chance to enter B. Of course this algorithm would be terrible from a practical perspective–the total number of particles is certain to grow exponentially. However, the algorithm is used only for the purposes of theoretical analysis, and the number of particles is not a concern. Overbars are used to distinguish this algorithm from the previous one.

Splitting Algorithm Fully Branching (SFB)

`Variables:`

$\bar{N}_r^n$ `number of particles in generation` r

$\bar{X}_{r,k}^n$ `position of` k^{th} `particle in generation` r

$\bar{w}_{r,k}^n$ `weight of` k^{th} `particle in generation` r

```
Initialization Step:
   $\bar{N}_0^n = 1$, $\bar{X}_{0,1}^n = x_n$, $\bar{w}_{0,1}^n = 1$
   for $r = 1, \ldots, l^n(x_n)$
      $\bar{N}_r^n = 0$
   end
Main Algorithm:
   for $r = 1, \ldots, l^n(x_n)$
      if $\bar{N}_{r-1}^n = 0$ then $\bar{N}_r^n = 0$
      else
         for $j = 1, \ldots, \bar{N}_{r-1}^n$
            generate $\bar{Z}_{r,j,i}^n$ a single sample of a process with the
            same law as $\bar{X}_i^n$ and initial condition $\bar{Z}_{r,j,0}^n = \bar{X}_{r-1,j}^n$

            let $\bar{\tau}_{r,j}^n = \min\{i : \bar{Z}_{r,j,i}^n \in A \cup C_{l^n(x_n)-r}^n\}$

            Splitting Step begin
            let $\bar{M}$ be an independent sample from the law $q$
            for $k = 1, \ldots, |r(\bar{M})|$
               $\bar{N}_r^n = \bar{N}_r^n + 1$
               $\bar{X}_{r,\bar{N}_r^n}^n = \bar{Z}_{r,j,\bar{\tau}_{r,j}^n}^n$
               $\bar{w}_{r,\bar{N}_r^n}^n = w_k(\bar{M})\bar{w}_{r-1,j}^n$
            end
            Splitting Step end

         end
      end
   end
Construction of a sample:
   once all the generations have been calculated we form
   the quantity
      $s_{\mathtt{SFB}}^n = \sum_{j=1}^{\bar{N}_{l^n(x_n)}^n} 1_{\left\{\bar{X}_{l^n(x_n),j}^n \in B\right\}} \bar{w}_{l^n(x_n),j}^n$.
```

Since the distributions of the two estimates coincide

$$E_{x_n}\left[\sum_{j=1}^{N_{l^n(x_n)}^n} w_{l^n(x_n),j}^n\right] = E_{x_n}\left[\sum_{j=1}^{\bar{N}_{l^n(x_n)}^n} 1_{\left\{\bar{X}_{l^n(x_n),j}^n \in B\right\}} \bar{w}_{l^n(x_n),j}^n\right].$$

Because of this, the SFB algorithm can be used to prove the following.

Lemma 5 *An estimator based on independent copies of s^n_{SA} is unbiased if and only if*

$$E\left[\sum_{i=1}^{r(M)} w_i(M)\right] = 1.$$

Proof. It suffices to prove

$$E_{x_n}\left[\sum_{j=1}^{\bar{N}^n_{l^n(x_n)}} 1_{\left\{\bar{X}^n_{l^n(x_n),j}\in B\right\}} \bar{w}^n_{l^n(x_n),j}\right] = p^n(x_n).$$

We will use a particular construction of the SFB algorithm that is useful here and elsewhere in the paper. Recall that with this algorithm every particle splits at every generation. Hence the random number of particles associated with each splitting can be generated prior to the generation of any trajectories that will determine particle locations. As a consequence, the total number of particles present at the last generation can be calculated, as can the weight that will be assigned to each particle in this final generation, prior to the assignment of a trajectory to the particle. Once the weights have been assigned, the trajectories of all the particles can be constructed in terms of random variables that are independent of those used to construct the weights. Since the probability that any such trajectory makes it to B prior to hitting A is $p^n(x_n)$,

$$E_{x_n}\left[\sum_{j=1}^{\bar{N}^n_{l^n(x_n)}} 1_{\left\{\bar{X}^n_{l^n(x_n),j}\in B\right\}} \bar{w}^n_{l^n(x_n),j}\right] = p^n(x_n) E_{x_n}\left[\sum_{j=1}^{\bar{N}^n_{l^n(x_n)}} \bar{w}^n_{l^n(x_n),j}\right].$$

A simple proof by induction and the independence of the splitting from particle to particle shows that

$$E_{x_n}\left[\sum_{j=1}^{\bar{N}^n_{l^n(x_n)}} \bar{w}^n_{l^n(x_n),j}\right] = \left(E\left[\sum_{i=1}^{r(M)} w_i(M)\right]\right)^{l^n(x_n)}.$$

■

For the rest of the paper we restrict attention to splitting mechanisms that are unbiased.

3 Stability

Now let an importance function V, level Δ, and splitting mechanism (q, r, w) be given. Define

$$\mathcal{J}(x, y) \doteq \inf_{\phi, t: \phi(0)=x, \phi(t)=y} \int_0^t L\left(\phi(s), \dot{\phi}(s)\right) ds, \tag{1}$$

where the infimum is over absolutely continuous functions. A function $\bar{W}$: $D \to \mathbb{R}$ will be called a *subsolution* if $\bar{W}(x) \leq 0$ for all $x \in B$ and if $\bar{W}(x) - \bar{W}(y) \leq \mathcal{J}(x, y)$ for all $x, y \in D \backslash (A \cup B)$. In Section 5 we will discuss conditions under which $\bar{W}$ can be identified as a viscosity subsolution for an associated PDE. Recall that a splitting algorithm is called stable if the total number of particles ever used grows subexponentially as $n \to \infty$. For a given splitting algorithm define

$$\bar{W}(x) = \frac{\log Er(M)}{\Delta} V(x). \tag{2}$$

In this section we show that, loosely speaking, a splitting algorithm is stable if and only if $\bar{W}$ is a subsolution.

A construction that will simplify some of the proofs is to replace a given splitting mechanism by one for which all the weights are constant. Thus (q, r, w) is replaced by $(q, r, \bar{w})$, where for each $j = 1, \ldots, J$ and $i = 1, \ldots, r(j)$,

$$[\bar{w}_i(j)]^{-1} = Er(M) = \sum_{j=1}^{J} r(j) q_j.$$

The new splitting mechanism is also unbiased, and the distribution of the number of particles at each stage is the same as that of (q, r, w).

To establish the instability we make a very mild assumption on a large deviation lower bound for the probability that an open ball is hit prior to reaching A. This assumption can be expected to hold under conditions which guarantee Condition 1.

Proposition 6 *Consider an importance function V, level Δ, and splitting mechanism (q, r, w), and define $\bar{W}$ by (2). Suppose there exists $y \in D \backslash (A \cup B)$ such that $\bar{W}(y) > 0$ and*

$$\bar{W}(x) - \bar{W}(y) > \mathcal{J}(x, y). \tag{3}$$

Assume that $\mathcal{J}(x,y)$ is continuous at y. Let $\tilde{p}^n(x_n)$ be the probability that X^n enters the ball of radius $\delta > 0$ about y before entering A, given $X_0^n = x_n$, and assume

$$\liminf_{n\to\infty} \frac{1}{n} \log \tilde{p}^n(x_n) \geq - \inf_{z:|z-y|<\delta} \mathcal{J}(x,z).$$

Then the corresponding splitting algorithm is not stable.

Proof. It is enough to prove the instability of the algorithm that uses $(q, r, \bar{w})$. Since $\mathcal{J}(x,y) > 0$, $V(y) < V(x)$. From the definition of $\bar{W}$ in (2) and (3) there exist $\delta > 0$ and $\varepsilon > 0$ such that for all z with $|y - z| \leq \delta$,

$$[V(x) - V(z)] \frac{\log Er(M)}{\Delta} > \mathcal{J}(x,z) + \varepsilon.$$

Let $S \doteq \{z : |y - z| < \delta\}$. By taking $\delta > 0$ smaller if necessary we can guarantee that $S \cap A = \emptyset$ and $V(z) > 0$ for all $z \in S$.

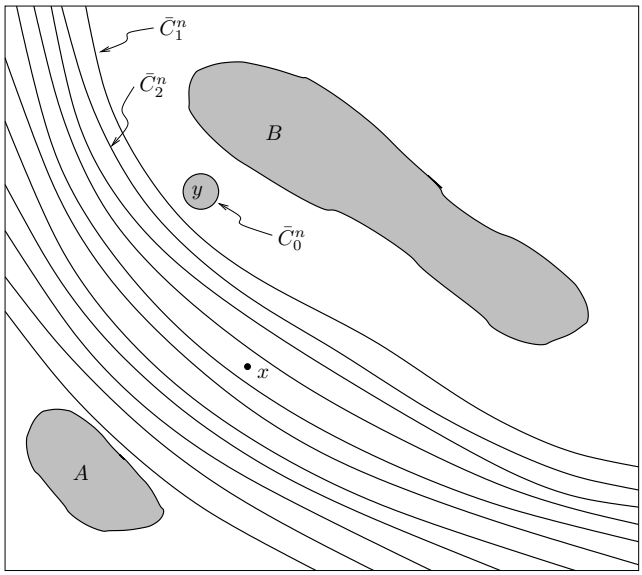

Figure 2: Level Sets of $\bar{V}$ in Proof of Instability.

Suppose one were to consider the problem of estimating $\tilde{p}^n(x_n)$. One could use the same splitting mechanism and level sets, and even the same random variables, except that one would stop on hitting A or S rather

than A or B, and the last stage would correspond to some number $m^n \leq l^n(x_n)$. Of course, since V is positive on S it will no longer work as an importance function, but there is a function $\bar{V} = V - a$ that will induce exactly the same level sets as V and which can serve as an importance function for this problem. See Figure 2. The two problems can be coupled, in that exactly the same random variables can be used to construct both the splitting mechanisms and particle trajectories up until the particles in the $\tilde{p}^n(x_n)$ problem enter $\bar{C}_1^n$.

If a particular particle has not been trapped in A prior to entering $\bar{C}_1^n$, then that particle would also not yet be trapped in A in the corresponding scheme used to estimate $p^n(x_n)$. Note also that the number of particles that make it to S are at most R times the number that make it to $\bar{C}_1^n$. Let $\tilde{N}_{m^n}^n$ denote the number of particles that make it to S in the SA used to approximate $\tilde{p}^n(x_n)$, and let $N_{l^n(x_n)}^n$ be the number used in the SA that approximates $p^n(x_n)$. Then $N_{l^n(x_n)}^n \geq \tilde{N}_{m^n}^n / R$.

Using the SFB variant in the same way that it was used in the proof of Lemma 5 and that the mechanism $(q, r, \bar{w})$ is used,

$$\begin{aligned}
\tilde{p}^n(x_n) &= E_{x_n}\left[\sum_{j=1}^{\tilde{N}_{m^n}^n} \tilde{w}_{m^n,j}^n\right] \\
&= E_{x_n}\left[\sum_{j=1}^{\tilde{N}_{m^n}^n} [Er(M)]^{-m^n}\right] \\
&= [Er(M)]^{-m^n} E_x\left[\tilde{N}_{m^n}^n\right].
\end{aligned}$$

We now use the lower bound on $\tilde{p}^n(x_n)$ and that $m^n/n \to [V(x) - \sup_{z\in S} V(z)]/\Delta$:

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \log E_{x_n}\left[\tilde{N}_{m^n}^n\right] \\
&= \liminf_{n\to\infty} \frac{1}{n} \log E_{x_n}\left[\tilde{p}^n(x_n)\,[Er(M)]^{m^n}\right] \\
&\geq -\inf_{z\in S} \mathcal{J}(x,z) + \frac{[V(x) - \sup_{z\in S} V(z)]}{\Delta} \log[Er(M)] \\
&\geq \inf_{z\in S}\left[\frac{[V(x) - V(z)]}{\Delta} \log[Er(M)] - \mathcal{J}(x,z)\right] \\
&\geq \varepsilon.
\end{aligned}$$

It follows that

$$\liminf_{n\to\infty} \frac{1}{n} \log E_{x_n}\left[N_{l^n(x_n)}^n\right] \geq \varepsilon > 0,$$

which completes the proof. ■

The next proposition considers stability. Here we will make a mild assumption concerning a large deviation upper bound, which can also be expected to hold under conditions which guarantee Condition 1.

Proposition 7 *Consider an importance function V, level Δ, and splitting mechanism (q, r, w), and define $\bar{W}$ by (2). Suppose that*

$$\bar{W}(x) - \bar{W}(y) \leq \mathcal{J}(x, y)$$

for all $x, y \in D \backslash (A \cup B)$ and that $\bar{W}(y) \leq 0$ for all $y \in B$. Consider any $a \in (0, V(x)/\Delta]$, let $\tilde{p}^n(x_n)$ be the probability that X^n enters level set L_a before entering A (given $X_0^n = x_n$), and assume

$$\limsup_{n\to\infty} \frac{1}{n} \log \tilde{p}^n(x_n) \leq - \inf_{z \in L_a} \mathcal{J}(x, z).$$

Then the corresponding splitting algorithm is stable.

Proof. For each n let r^n be the value in $\{1, \ldots, l^n(x)\}$ that maximizes $r \to E_x[N_r^n]$. Since r^n/n is bounded, along some subsequence (again denoted by n) we have $r^n/n \to v \in [0, V(x)/\Delta]$. Using the usual argument by contradiction, it is enough to prove

$$\limsup_{n\to\infty} \frac{1}{n} \log E_{x_n}[N_{r^n}^n] \leq 0$$

along this subsequence. First suppose that $v = 0$. Given $\delta > 0$, choose $\bar{n} < \infty$ such that $r^n/n \leq \delta$ for all $n \geq \bar{n}$. Then $N_{r^n}^{n,x_n} \leq R^{\delta n}$, and so $\limsup_{n\to\infty} \frac{1}{n} \log E_{x_n}[N_{r^n}^{n,x_n}] \leq \delta \cdot \log R$. Since $\delta > 0$ is arbitrary, this case is complete.

Now assume $v \in (0, V(x)/\Delta]$ and let $\delta \in (0, v)$ be given. Suppose one were to consider the problem of estimating $\tilde{p}^n(x_n)$ as defined in the statement of the proposition, with $a = v$. We again use the same splitting mechanism and level sets, except that we now stop on hitting A or $L_{v+\delta}$. An importance function with these level sets can be found by adding a constant to V. We again couple the processes, and observe that entry into $\bar{C}_1^n$ for the $\tilde{p}^n(x_n)$ problem corresponds to entry into $C_{m_n}^n$ in the $p^n(x_n)$ problem, where $m_n/n \to (V(x) - v - \delta)/\Delta$ as $n \to \infty$. Observe that every particle in the

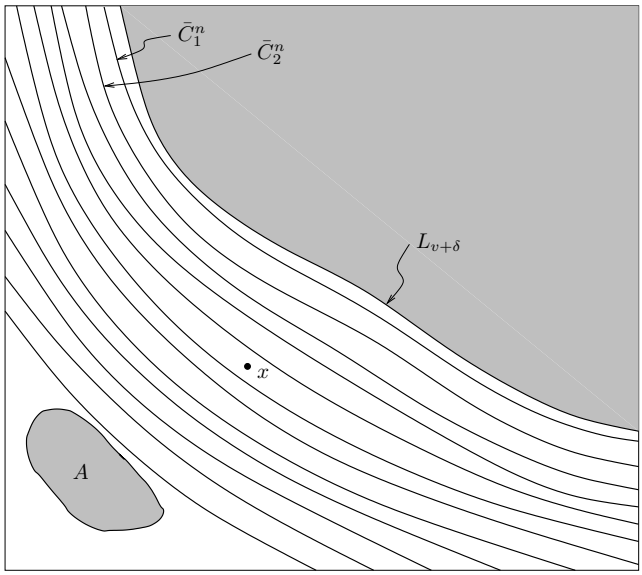

Figure 3: Level Sets of $\bar{V}$ in Proof of Stability.

algorithm used to estimate $p^n(x_n)$ that is not trapped in A by stage m^n is also not trapped in A in the algorithm used to estimate $\tilde{p}^n(x_n)$. Hence the number of such particles can serve as an upper bound on the number used to construct s^n_{SA}. Let $\tilde{N}^n_{m^n}$ denote the number of such particles for the SA used to estimate $\tilde{p}^n(x_n)$.

We again use the SFB variant in the same way that it was used in the proof of Lemma 5 and the $(q, r, \bar{w})$ splitting mechanism to obtain

$$\tilde{p}^n(x_n) = [Er(M)]^{-m^n} E_{x_n}\left[\tilde{N}^n_{m^n}\right].$$

Using the upper bound on $\tilde{p}^n(x_n)$ and that $m^n/n \to [V(x) - v - \delta]/\Delta$:

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log E_{x_n}\left[\tilde{N}^n_{m^n}\right] &= \limsup_{n\to\infty} \frac{1}{n} \log E_{x_n}\left[\tilde{p}^n(x_n)\,[Er(M)]^{m^n}\right] \\
&\leq -\inf_{z\in L_{v+\delta}} \mathcal{J}(x,z) + \frac{[V(x) - v - \delta]}{\Delta} \log[Er(M)] \\
&= \sup_{z\in L_{v+\delta}} \left[\frac{[V(x) - V(z)]}{\Delta} \log[Er(M)] - \mathcal{J}(x,z)\right] \\
&\leq 0.
\end{aligned}$$

For sufficiently large n we have $r_n - m_n \leq 2\delta n/\Delta$, and hence $N^n_{r^n} \leq \tilde{N}^n_{m^n}$ ·

$R^{2\delta n/\Delta}$. It follows that

$$\limsup_{n\to\infty} \frac{1}{n} \log E_{x_n} [N^n_{r^n}] \leq (2\delta/\Delta) \cdot \log R,$$

and since $\delta > 0$ is arbitrary the proof is complete. ■

4 Asymptotic Performance

Since the sample s^n_{SA} has mean $p^n(x_n)$, any estimator constructed as an average of independent copies of s^n_{SA} is unbiased and has variance proportional to $\mathrm{var}_{x_n} [s^n_{\mathrm{SA}}]$. Once the mean is fixed, the minimization of $\mathrm{var}_{x_n} [s^n_{\mathrm{SA}}]$ among splitting algorithms is equivalent to the minimization of $E_{x_n} [s^n_{\mathrm{SA}}]^2$. It is of course very difficult to find the minimizer in this problem. When a large deviation scaling holds, a useful alternative is to maximize the rate of decay of the second moment, i.e., to maximize

$$\liminf_{n\to\infty} -\frac{1}{n} \log E_{x_n} [s^n_{\mathrm{SA}}]^2 = \liminf_{n\to\infty} -\frac{1}{n} \log E_{x_n} \left[\sum_{j=1}^{N^n_{l^n(x_n)}} w^n_{l^n(x_n),j} \right]^2 .$$

By Jensen's inequality the best possible rate is $2W(x)$:

$$\liminf_{n\to\infty} -\frac{1}{n} \log E_{x_n} [s^n_{\mathrm{SA}}]^2 \geq \liminf_{n\to\infty} -\frac{2}{n} \log E_{x_n} [s^n_{\mathrm{SA}}] \geq 2W(x).$$

The main result of this section is the following.

Theorem 8 *Consider an importance function V, level Δ, and splitting mechanism (q, r, w), and define $\bar{W}$ by (2). Suppose that*

$$\bar{W}(x) - \bar{W}(y) \leq \mathcal{J}(x, y)$$

for all $x, y \in D \setminus (A \cup B)$ and that $\bar{W}(y) \leq 0$ for all $y \in B$. Assume also that Conditions 1 and 4 hold. Then

$$\lim_{n\to\infty} -\frac{1}{n} \log E_{x_n} [s^n_{\mathrm{SA}}]^2 = W(x) - V(x) \frac{\log \left(E \sum_{i=1}^{r(M)} w_i(M)^2 \right)}{\Delta}.$$

Proof. It is sufficient to consider the SFB algorithm and prove that

$$\begin{aligned}\lim_{n\to\infty} -\frac{1}{n}\log E_{x_n}\left[\sum_{j=1}^{\bar{N}^n_{l^n(x_n)}} 1_{\left\{\bar{X}^n_{l^n(x_n),j}\in B\right\}}\bar{w}^n_{l^n(x_n),j}\right]^2 \\ = \quad W(x) - V(x)\frac{\log\left(E\sum_{i=1}^{r(M)} w_i(M)^2\right)}{\Delta}.\end{aligned}$$

The proof is broken into upper and lower bounds.

We first prove

$$\begin{aligned}\limsup_{n\to\infty} -\frac{1}{n}\log E_{x_n}\left[\sum_{j=1}^{\bar{N}^n_{l^n(x_n)}} 1_{\left\{\bar{X}^n_{l^n(x_n),j}\in B\right\}}\bar{w}^n_{l^n(x_n),j}\right]^2 \\ \leq \quad W(x) - V(x)\frac{\log\left(E\sum_{i=1}^{r(M)} w_i(M)^2\right)}{\Delta}.\end{aligned} \tag{4}$$

In the following display we drop cross terms to obtain the inequality, and then use the same construction as in Lemma 5 under which the weights and trajectories are independent to obtain the equality.

$$\begin{aligned}&\limsup_{n\to\infty} -\frac{1}{n}\log E_{x_n}\left[\sum_{j=1}^{\bar{N}^n_{l^n(x_n)}} 1_{\left\{\bar{X}^n_{l^n(x_n),j}\in B\right\}}\bar{w}^n_{l^n(x_n),j}\right]^2 \\ &\leq \quad \limsup_{n\to\infty} -\frac{1}{n}\log E_{x_n}\left[\sum_{j=1}^{\bar{N}^n_{l^n(x_n)}} 1_{\left\{\bar{X}^n_{l^n(x_n),j}\in B\right\}}\left(\bar{w}^n_{l^n(x_n),j}\right)^2\right] \\ &= \quad \limsup_{n\to\infty} -\frac{1}{n}\log\left(p^n(x_n)E_{x_n}\left[\sum_{j=1}^{\bar{N}^n_{l^n(x_n)}}\left(\bar{w}^n_{l^n(x_n),j}\right)^2\right]\right).\end{aligned}$$

Suppose we prove that for any κ (and in particular $\kappa = l^n(x_n)$), that

$$E_{x_n}\left[\sum_{j=1}^{\bar{N}^n_\kappa}\left(\bar{w}^n_{\kappa,j}\right)^2\right] = \left(E\sum_{i=1}^{r(M)} w_i(M)^2\right)^\kappa. \tag{5}$$

Since $l^n(x_n) = \lceil nV(x_n)/\Delta\rceil$, (4) will follow from Condition 1. The proof of (5) is by induction. Let M_j denote the independent random variables used

to define the splitting for the jth particle at stage κ. Then

$$\begin{aligned} E_{x_n}\left[\sum_{j=1}^{\bar{N}_{\kappa+1}^n}\left(\bar{w}_{\kappa+1,j}^n\right)^2\right] &= E_{x_n}\left[\sum_{j=1}^{\bar{N}_\kappa^n}\left(\bar{w}_{\kappa,j}^n\right)^2\sum_{i=1}^{r(M_j)} w_i(M_j)^2\right] \\ &= E_{x_n}\left[\sum_{j=1}^{\bar{N}_\kappa^n}\left(\bar{w}_{\kappa,j}^n\right)^2\right]\left(E\sum_{i=1}^{r(M)} w_i(M)^2\right) \\ &= \left(E\sum_{i=1}^{r(M)} w_i(M)^2\right)^{\kappa+1}. \end{aligned}$$

We now turn to the proof of the lower bound

$$\begin{aligned} &\liminf_{n\to\infty} -\frac{1}{n}\log E_{x_n}\left[\sum_{j=1}^{\bar{N}_{l^n(x_n)}^n} 1_{\left\{\bar{X}_{l^n(x_n),j}^n\in B\right\}}\bar{w}_{l^n(x_n),j}^n\right]^2 \qquad (6) \\ &\geq W(x) - V(x)\frac{\log\left(E\sum_{i=1}^{r(M)} w_i(M)^2\right)}{\Delta}. \end{aligned}$$

For each stage κ, let $M_{\kappa,j}^n$ denote the independent random variables used in the splitting of particle $j \in \left\{1,\ldots,\bar{N}_\kappa^n\right\}$. Also, let $I_{\kappa,j}^n$ denote the disjoint decomposition of the particles in $\left\{1,\ldots,\bar{N}_{\kappa+1}^n\right\}$ according to their parent particle. Observe that if $k,l \in I_{\kappa,j}^n, k \neq l$, then for all particles descended from k and l, κ is the time of their last common ancestor. Given $k \in I_{\kappa,j}^n$, let $\bar{I}_{\kappa+1,l^n(x_n),k}^n$ denote the descendants of this particle at stage $l^n(x_n)$. With

this notation we can write

$$
\begin{aligned}
& E_{x_n}\left[\sum_{j=1}^{\bar{N}^n_{l^n(x_n)}} 1_{\left\{\bar{X}^n_{l^n(x_n),j}\in B\right\}}\bar{w}^n_{l^n(x_n),j}\right]^2 \\
&= \sum_{\kappa=1}^{l^n(x_n)-1} E_{x_n}\left[\sum_{j=1}^{\bar{N}^n_\kappa}\sum_{k,l\in I^n_{\kappa,j},k\neq l}\sum_{m_k\in\bar{I}^n_{\kappa+1,l^n(x_n),k}} 1_{\left\{\bar{X}^n_{l^n(x_n),m_k}\in B\right\}}\bar{w}^n_{l^n(x_n),m_k}\right. \\
&\qquad\left.\cdot\sum_{m_l\in\bar{I}^n_{\kappa+1,l^n(x_n),k}} 1_{\left\{\bar{X}^n_{l^n(x_n),m_l}\in B\right\}}\bar{w}^n_{l^n(x_n),m_l}\right] \\
&\quad + E_{x_n}\left[\sum_{j=1}^{\bar{N}^n_{l^n(x_n)}} 1_{\left\{\bar{X}^n_{l^n(x_n),j}\in B\right\}}\left(\bar{w}^n_{l^n(x_n),j}\right)^2\right].
\end{aligned}
$$

Let $\bar{w}^n_{\kappa,j}$ denote the products of the weights accumulated by particle $j\in\left\{1,\ldots,\bar{N}^n_\kappa\right\}$ up to stage κ, and let $\bar{w}^n_{\kappa+1,l^n(x_n),m}$ denote the product of the weights accumulated by particle $m\in\left\{1,\ldots,\bar{N}^n_{l^n(x_n)}\right\}$ between stages $\kappa+1$ and the final stage. Finally, let $\mathcal{F}^n_\kappa$ denote the sigma algebra generated by $M^n_{s,j}, s\in\{1,\ldots,\kappa\}, j\in\left\{1,\ldots,\bar{N}^n_s\right\}$ and the random variables used to construct $\bar{X}^n_{s,j}$ for these same indices. Note that the future weights are independent of $\mathcal{F}^n_\kappa$, and that the distribution of $\bar{X}^n_{l^n(x_n),m}$ depends on $\mathcal{F}^n_\kappa$ only through $\bar{X}^n_{\kappa,j}$ if $k\in I^n_{\kappa,j}$ and $m\in\bar{I}^n_{\kappa+1,l^n(x_n),k}$. We introduce the notation

$$
\begin{aligned}
Y^n_{\kappa,j} &\doteq 1_{\left\{\bar{X}^n_{\kappa,j}\notin A\right\}}\left(\bar{w}^n_{\kappa,j}\right)^2, \\
Z^n_{\kappa,k} &\doteq \sum_{m\in\bar{I}^n_{\kappa+1,l^n(x_n),k}} 1_{\left\{\bar{X}^n_{l^n(x_n),m}\in B\right\}}\bar{w}^n_{\kappa+1,l^n(x_n),m}
\end{aligned}
$$

By conditioning on $\mathcal{F}_\kappa^n$ we get

$$E_{x_n}\left[\sum_{j=1}^{\bar{N}_\kappa^n}\sum_{k,l\in I_{\kappa,j}^n,k\neq l}\sum_{m_k\in\bar{I}_{\kappa+1,l^n(x_n),k}^n} 1_{\left\{\bar{X}_{l^n(x_n),m_k}^n\in B\right\}}\bar{w}_{l^n(x_n),m_k}^n \right. \tag{7}$$

$$\left. \cdot \sum_{m_l\in\bar{I}_{\kappa+1,l^n(x_n),k}^n} 1_{\left\{\bar{X}_{l^n(x_n),m_l}^n\in B\right\}}\bar{w}_{l^n(x_n),m_l}^n\right]$$

$$= E_{x_n}\left[\sum_{j=1}^{\bar{N}_\kappa^n} Y_{\kappa,j}^n \sum_{k,l\in I_{\kappa,j}^n,k\neq l} w_k(M_{\kappa,j}^n)Z_{\kappa,k}^n w_l(M_{\kappa,j}^n)Z_{\kappa,l}^n\right]$$

$$= E_{x_n}\left[\sum_{j=1}^{\bar{N}_\kappa^n} Y_{\kappa,j}^n \sum_{k,l\in I_{\kappa,j}^n,k\neq l} w_k(M_{\kappa,j}^n)w_l(M_{\kappa,j}^n)E_{\bar{X}_{\kappa,j}^n}\left[Z_{\kappa,k}^n\right]E_{\bar{X}_{\kappa,j}^n}\left[Z_{\kappa,l}^n\right]\right].$$

Using again the independence of the weights and trajectories as used in the proof of Lemma 5, we have

$$E_{\bar{X}_{\kappa,j}^n}\left[Z_{\kappa,k}^n\right] = p^n\left(\bar{X}_{\kappa,j}^n\right).$$

Since

$$\mathcal{W} \doteq E\sum_{k\neq l} w_k(M)w_l(M) = E\left[\sum_k w_k(M)\right]^2 - E\left[\sum_k w_k(M)^2\right],$$

the final expression in (7) equals

$$\mathcal{W}E_{x_n}\left[\sum_{j=1}^{\bar{N}_\kappa^n} Y_{\kappa,j}^n p^n\left(\bar{X}_{\kappa,j}^n\right)^2\right].$$

We conclude that

$$E_{x_n}\left[\sum_{j=1}^{\bar{N}_{l^n(x_n)}^n} 1_{\left\{\bar{X}_{l^n(x_n),j}^n\in B\right\}}\bar{w}_{l^n(x_n),j}^n\right]^2$$

$$= \mathcal{W}\sum_{\kappa=1}^{l^n(x_n)-1} E_{x_n}\left[\sum_{j=1}^{\bar{N}_\kappa^n} Y_{\kappa,j}^n p^n\left(\bar{X}_{\kappa,j}^n\right)^2\right]$$

$$+E_{x_n}\left[\sum_{j=1}^{\bar{N}_{l^n(x_n)}^n} 1_{\left\{\bar{X}_{l^n(x_n),j}^n\in B\right\}}\left(\bar{w}_{l^n(x_n),j}^n\right)^2\right].$$

For a final time we use that the weights and trajectories are independent, and also (5), to argue that for any bounded and measurable function F and any stage κ,

$$\begin{aligned} E_{x_n}\left[\sum_{j=1}^{\bar{N}_\kappa^n} F\left(\bar{X}_{\kappa,j}^n\right)\left(\bar{w}_{\kappa,j}^n\right)^2\right] &= E_{x_n}\left[F\left(\bar{X}_{\kappa,1}^n\right)\right] E_{x_n}\left[\sum_{j=1}^{\bar{N}_\kappa^n}\left(\bar{w}_{\kappa,j}^n\right)^2\right] \\ &= E\left[\sum_{i=1}^{r(M)} w_i(M)^2\right]^\kappa E_{x_n}\left[F\left(\bar{X}_{\kappa,1}^n\right)\right]. \end{aligned}$$

Thus

$$\begin{aligned} &E_{x_n}\left[\sum_{j=1}^{\bar{N}_{l^n(x_n)}^n} 1_{\left\{\bar{X}_{l^n(x_n),j}^n \in B\right\}} \bar{w}_{l^n(x_n),j}^n\right]^2 \\ &= \mathcal{W}\sum_{\kappa=1}^{l^n(x_n)-1}\left(E\sum_{i=1}^{r(M)} w_i(M)^2\right)^\kappa E_{x_n}\left[1_{\left\{\bar{X}_{\kappa,1}^n \notin A\right\}} p^n\left(\bar{X}_{\kappa,1}^n\right)^2\right] \\ &\quad + \left(E\sum_{i=1}^{r(M)} w_i(M)^2\right)^{l^n(x_n)} E_{x_n}\left[1_{\left\{\bar{X}_{l^n(x_n),1}^n \notin A\right\}}\right]. \end{aligned}$$

Since $l^n(x_n)$ is proportional to n, to prove (6) it is enough to show that if κ_n is any sequence such that $\kappa_n/n \to v \in [0, V(x)/\Delta]$, then

$$\begin{aligned} &\liminf_{n\to\infty} -\frac{1}{n}\log E_{x_n}\left[\left(E\sum_{i=1}^{r(M)} w_i(M)^2\right)^{\kappa_n} E_{x_n}\left[1_{\left\{\bar{X}_{\kappa_n,1}^n \notin A\right\}} p^n\left(\bar{X}_{\kappa_n,1}^n\right)^2\right]\right] \\ &\qquad \geq W(x) - V(x)\frac{\log\left(E\sum_{i=1}^{r(M)} w_i(M)^2\right)}{\Delta}. \end{aligned}$$

Observe that $\left\{\bar{X}_{\kappa_n,1}^n \notin A\right\}$ implies $\bar{X}_{\kappa_n,1}^n \in C_{\lceil nV(x)/\Delta\rceil-\kappa_n}^n$. By Condition 4,

$$\liminf_{n\to\infty} -\frac{1}{n}\log E_{x_n}\left[1_{\left\{\bar{X}_{\kappa_n,1}^n \notin A\right\}} p^n\left(\bar{X}_{\kappa_n,1}^n\right)^2\right] \geq W(x) + \inf_{y\in\partial L_{V(x)-v\Delta}} W(y).$$

By the subsolution property, $W(y) \geq V(y)\log Er(M)/\Delta$. Since Holder's

inequality gives $-\log\left(E\sum_{i=1}^{r(M)} w_i(M)^2\right) \leq \log\left(Er(M)\right)$,

$$\begin{aligned}
&\liminf_{n\to\infty} -\frac{1}{n}\log E_{x_n}\left[\left(E\sum_{i=1}^{r(M)} w_i(M)^2\right)^{\kappa_n} E_{x_n}\left[1_{\{\bar{X}^n_{\kappa_n,1}\notin A\}} p^n\left(\bar{X}^n_{\kappa_n,1}\right)^2\right]\right] \\
&\quad\geq\quad -v\log\left(E\sum_{i=1}^{r(M)} w_i(M)^2\right) + W(x) + \inf_{y\in\partial L_{V(x)-v\Delta}} \frac{\log Er(M)}{\Delta} V(y) \\
&\quad=\quad -v\log\left(E\sum_{i=1}^{r(M)} w_i(M)^2\right) + W(x) + \log Er(M)\left(\frac{V(x)}{\Delta} - v\right) \\
&\quad\geq\quad W(x) - V(x)\frac{\log\left(E\sum_{i=1}^{r(M)} w_i(M)^2\right)}{\Delta},
\end{aligned}$$

and the proof is complete. ■

4.1 Design of a Splitting Algorithm

Suppose that $V(x)$ and Δ are given and that we choose a splitting mechanism (q, r, w) which is unbiased and stable. By Theorem 8 the asymptotic rate of decay of the second moment is given by

$$W(x) - V(x)\frac{\log\left(E\sum_{i=1}^{r(M)} w_i(M)^2\right)}{\Delta}.$$

By Hölder's inequality

$$E\sum_{i=1}^{r(M)} w_i(M)^2 \cdot Er(M) \geq E\sum_{i=1}^{r(M)} w_i(M) = 1,$$

and therefore

$$-\log\left(E\sum_{i=1}^{r(M)} w_i(M)^2\right) \leq \log\left(Er(M)\right).$$

Equality holds if and only if $w_i(m) = \frac{1}{Er(M)}$ for all $i \in \{1, \ldots, r(m)\}$ and all $m \in \{1, \ldots, J\}$. Given the value $u = Er(M)$, an alternative splitting

mechanism which is arguably the simplest which preserves the value and achieves the equality in Holder's inequality is that defined by $J = 2$ and

$$q_1 = \lceil u \rceil - u, q_2 = 1 - q_1,\ r(1) = \lfloor u \rfloor, r(1) = \lfloor u \rfloor + 1,\ w_i(j) = 1/u \text{ all } i, j. \tag{8}$$

Given a subsolution $\bar{W}$, the design problem and the performance of the resulting algorithm can be summarized as follows.

- Choose a level Δ and mean number of particles u, and define an importance function V by $\log u \cdot V(x)/\Delta = \bar{W}(x)$. Define the splitting mechanism by (8). The resulting splitting algorithm will be stable.
- If s^n_{SA} is a single sample constructed according to this algorithm, then we have the asymptotic performance

$$\lim_{n\to\infty} -\frac{1}{n} \log E_{x_n}[(s^n_{\mathrm{SA}})^2] = W(x) + \bar{W}(x).$$

- The largest possible subsolution satisfies $\bar{W}(x) = W(x)$, in which case we achieve asymptotically optimal performance.

Remark 9 Although the subsolution property guarantees stability, it could allow for polynomial growth of the number of particles. If in practice one observes that a large number of particles make it to B in the course of simulating a single sample s^n_{SA}, then one can consider reducing the value of Δ slightly, while keeping the mechanism and V fixed. This will increase the second moment of the estimator slightly, but will also lead to an algorithm that requires little effort for each single sample.

5 The Associated Hamilton-Jacobi-Bellman Equation

The probability $p^n(x)$ is intimately and naturally related, via the exponential rate $W(x)$, with a certain nonlinear PDE. This relation is well known, and follows from the fact that W is characterized in terms of an optimal control or calculus of variations problem. We begin this section by defining the PDE and the notion of a *subsolution* in the PDE context.

Our interest in this characterization is because it is more convenient for the explicit construction of subsolutions than the one based on the calculus of variations problem. See, for example, the subsolutions constructed for

various large deviation problems in [5] and [3]. (It should be noted that the constructions in these papers ultimately produce *classical* subsolutions. In contrast, the splitting algorithms require only the weaker viscosity subsolution property. However, the smoother subsolutions constructed in [3, 5] are obtained in as mollified versions of viscosity subsolutions, and it is the construction of these unmollified functions that is relevant to the present paper.) Other examples will be given in the next section. Since our only interest in the PDE is as a tool for explicit constructions, we describe the characterization formally and in the simplest possible setting, and refer the reader to [1, 6].

For $q \in \mathbb{R}^d$, let

$$\mathbb{H}(x,q) = \inf_{\beta \in \mathbb{R}^d} \left[\langle q, \beta \rangle + L(x,\beta) \right].$$

Then under regularity conditions on L and the sets A and B, W can be characterized as the maximal viscosity subsolution to

$$\mathbb{H}(x, D\bar{W}(x)) = 0, x \notin A \cup B, \bar{W}(x) = \begin{cases} 0 & x \in \partial B \\ \infty & x \in \partial A \end{cases}.$$

A continuous function $\bar{W}$ is a viscosity subsolution to this equation and boundary conditions if $\bar{W}(x) \leq 0$ for $x \in \partial B$, $\bar{W}(x) \leq \infty$ for $x \in \partial A$ and if the following condition holds. If $\phi : \mathbb{R}^d \to \mathbb{R}$ is a smooth test function such that the mapping $x \to \left[\bar{W}(x) - \phi(x)\right]$ attains a maximum at $x_0 \in \mathbb{R}^d \backslash (A \cup B)$, then $\mathbb{H}(x_0, D\phi(x_0)) \geq 0$.

Note that $\mathbb{H}(x, \cdot)$ is concave for each $x_0 \in \mathbb{R}^d$, and hence the pointwise minimum of a collection of subsolutions is again a subsolution. It is this observation which makes the explicit construction of subsolutions feasible in a number of interesting problems (see [5]).

In Section 2 we defined $\bar{W}(x)$ to be a subsolution to the calculus of variations problem if it satisfied the boundary inequalities and

$$\bar{W}(x) - \bar{W}(y) \leq \mathcal{J}(x,y)$$

for all $x, y \in \mathbb{R}^d \backslash (A \cup B)$. We now give the elementary proof that these notions coincide. Let $\bar{W}(x) - \phi(x)$ attain a maximum at x_0. Thus for any $\beta \in \mathbb{R}^d$ and all $a \in (0,1)$ sufficiently small, $\bar{W}(x_0 + a\beta) - \phi(x_0 + a\beta) \leq \bar{W}(x_0) - \phi(x_0)$, and so

$$\begin{aligned} \phi(x_0) - \phi(x_0 + a\beta) &\leq \bar{W}(x_0) - \bar{W}(x_0 + a\beta) \\ &\leq \mathcal{J}(x_0, x_0 + a\beta). \end{aligned}$$

Since $\mathcal{J}(x, y)$ is defined as an infimum over all trajectories that connect x to y, we always have

$$\mathcal{J}(x_0, x_0 + a\beta) \leq \int_0^a L(x_0 + s\beta, \beta) ds.$$

Hence if, e.g., the mapping $x \to L(x, \beta)$ is continuous, then for all β

$$\phi(x_0) - \phi(x_0 + a\beta) \leq L(x_0, \beta)a + o(a).$$

Using Taylor's Theorem to expand ϕ, sending $a \downarrow 0$ and then infimizing over β gives

$$0 \leq \inf_{\beta \in \mathbb{R}^d} [\langle D\phi(x_0), \beta \rangle + L(x_0, \beta)] \leq \mathbb{H}(x, D\phi(x_0)).$$

Thus $\bar{W}$ is a subsolution.

The calculation just given does not show that W is the maximal viscosity subsolution, or even that it is always safe to use a viscosity subsolution to the PDE in the design of a splitting scheme. The characterization of W as the *maximal* viscosity subsolution requires that we establish $\bar{W}(x) \leq W(x)$ whenever $\bar{W}$ is a viscosity subsolution. A standard approach to this would be to show that given a viscosity subsolution $\bar{W}$, any point $x \notin A \cup B$, and any $\varepsilon > 0$, there exists a smooth classical subsolution $\bar{W}^\varepsilon$ such that $\bar{W}^\varepsilon(x) \geq \bar{W}(x) - \varepsilon$. When this is true the classical verification argument [6] can be sued to show $W(x) \geq \bar{W}^\varepsilon(x)$, and since $\varepsilon > 0$ is arbitrary $W(x) \geq \bar{W}(x)$. This brings us very close to the method of constructing nearly optimal importance sampling schemes as described in [3, 5], where the design of the scheme must be based on the smooth classical subsolution $\bar{W}^\varepsilon(x)$ rather than $\bar{W}(x)$. In all the examples of the next subsection the inequality $\bar{W}(x) \leq W(x)$ can be established by constructing a nearby smooth subsolution as in [3, 5].

6 Numerical Examples

In this section we present some numerical results. We study four problems: buffer overflow for a tandem Jackson network with one shared buffer, simultaneous buffer overflow for a tandem Jackson network with separate buffers for each queue, some buffer overflow problems for a simple Markov modulated queue and estimation of the sample mean of a sequence of i.i.d. random variables.

Subsolutions, even among those with the maximal value at a given point, are not unique, and indeed for the problems to be discussed there are sometimes a number of reasonable choices one could make. We will not give

any details of the proof of the subsolution property, but simply note that in each case it can be proved by a direct verification argument as discussed in Section 5.

6.1 Tandem Jackson Network - Single Shared Buffer

Consider a stable tandem Jackson network with service rates $\lambda < \min\{\mu_1, \mu_2\}$. Suppose that the two queues share a single buffer and that we are interested in the probability

$$p^n = P_{(0,0)}\{\text{total population reaches } n \text{ before first return to } (0,0)\}$$

It is well known that

$$\lim_{n\to\infty} -\frac{1}{n}\log p^n = \min\{\rho_1, \rho_2\}$$

where $\rho_i = \log\frac{\mu_i}{\lambda}$. Further, the (continuous time) Hamiltonian that corresponds to subsolutions of the relevant calculus of variations problem is

$$\mathbb{H}(p) = -[\lambda(e^{-p_1} - 1) + \mu_1(e^{(p_1-p_2)} - 1) + \mu_2(e^{p_2} - 1)].$$

(see [3] for the discrete time analogue). Without loss of generality (see [3]) one can assume that $\mu_2 \le \mu_1$. By inspection $\mathbb{H}(p) = 0$ for $p = -\log\frac{\mu_2}{\lambda}(1, 1)$ (this root is suggested by the form of the escape region), and $W(x) = \langle p, x\rangle + \log\frac{\mu_2}{\lambda}$ is a subsolution which in fact takes the maximal value at $(0, 0)$ and so leads to an asymptotically optimal splitting scheme. The table below shows the results of a splitting simulation with 20,000 runs for $\lambda = 1$, $\mu_1 = \mu_2 = 4.5$ and for various values of n.

n	30	40	50
Theoretical Value	2.63×10^{-18}	1.03×10^{-24}	3.80×10^{-31}
Estimate	2.69×10^{-18}	0.97×10^{-24}	3.98×10^{-31}
Std. Err.	0.11×10^{-18}	0.04×10^{-24}	0.20×10^{-31}
95% C.I.	$[2.48, 2.90] \times 10^{-18}$	$[0.88, 1.05] \times 10^{-24}$	$[3.60, 4.37] \times 10^{-31}$
Time Taken (s)	32	67	165
Total no. successes	471776	579127	810382
Max no. particles	4027	4134	6987

Table 1. $\lambda = 1$, $\mu_1 = \mu_2 = 4.5$, asymptotically optimal scheme.

It was noted in Remark 9 that the number of particles generated may grow subexponentially in n and this appears to be reflected in the data. Following the suggestion of the remark, we also considered a slightly suboptimal subsolution in the hopes of better controlling the number of particles with little loss in performance. The table below shows the results of numerical simulation for the same problem with a splitting algorithm based on the

subsolution $\bar{W}(x) = \frac{4.2}{4.5}W(x)$. Again each estimate is obtained using 20,000 runs. The results are in accord with our expectations.

n	30	40	50
Theoretical Value	2.63×10^{-18}	1.03×10^{-24}	3.80×10^{-31}
Estimate	2.80×10^{-18}	1.10×10^{-24}	3.70×10^{-31}
Std. Err.	0.14×10^{-18}	0.08×10^{-24}	0.31×10^{-31}
95% C.I.	$[2.51, 3.10] \times 10^{-18}$	$[0.94, 1.26] \times 10^{-24}$	$[3.09, 4.31] \times 10^{-31}$
Time Taken (s)	6	10	15
Total no. successes	66607	44707	25578
Max no. particles	806	1145	678

Table 2. $\lambda = 1$, $\mu_1 = \mu_2 = 4.5$, asymptotically suboptimal scheme.

6.2 Tandem Jackson Network - Separate Buffers

In the paper [8] the authors address the problem of asymptotic optimality for splitting algorithms. In particular they consider an approach to choosing level sets that are claimed to be "consistent" with the large deviations analysis and show that this does not always lead to asymptotically optimal algorithms. They illustrate their results by considering, for a tandem Jackson network, the problem of simulating the probabilities

$$p^n = P_{(0,0)}\{\text{both queues simultaneously exceed } n \text{ before first return to } (0,0)\}.$$

It is shown that

$$\lim_{n\to\infty} -\frac{1}{n}\log p^n = \rho_1 + \rho_2 \doteq \gamma,$$

and the authors propose a splitting algorithm based on the importance function $U(x) = \gamma - \gamma\min\{x_1, x_2\}$, which is just a rescaling of the target set $B = \{(x, y) : x \geq n \text{ or } y \geq n\}$. They show that although the level sets given by this function may intuitively seem to agree with the most likely path to the rare set identified by the large deviations analysis, the resulting splitting algorithm in fact has very poor performance. By analyzing this importance function using the subsolution approach it is very easy to see why this is the case. The Hamiltonian corresponding to subsolutions is the same as in the previous section and it is clear to see that $U(x)$ is not a subsolution. However, the function

$$W(x) = \gamma - \rho_1 x_1 - \rho_2 x_2$$

is a subsolution. Further $W(0) = \gamma$, thus the corresponding importance function will lead to an asymptotically optimal splitting algorithm. Numerical results are presented for the cases $\lambda = 1, \mu_1 = 3, \mu_2 = 2$ and $\lambda = 1, \mu_1 = 2, \mu_2 = 3$ which are the same rates originally considered in

[8]. Each estimate was obtained by a simulation using 20,000 runs.

n	10	20	30
Theoretical Value	9.64×10^{-8}	1.60×10^{-15}	2.64×10^{-23}
Estimate	9.61×10^{-8}	1.60×10^{-15}	2.50×10^{-23}
Std. Err.	0.17×10^{-8}	0.03×10^{-15}	0.06×10^{-23}
95% C.I.	$[9.28, 9.94] \times 10^{-8}$	$[1.53, 1.66] \times 10^{-15}$	$[2.38, 2.61] \times 10^{-23}$
Time Taken (s)	17	137	454
Total no. successes	116255	116918	110337
Max no. particles	672	2274	3864

Table 3. $\lambda = 1, \mu_1 = 3, \mu_2 = 2$, asymptotically optimal scheme.

n	10	20	30
Theoretical Value	9.64×10^{-8}	1.60×10^{-15}	2.64×10^{-23}
Estimate	9.49×10^{-8}	1.67×10^{-15}	2.76×10^{-23}
Std. Err.	0.27×10^{-8}	0.07×10^{-15}	0.14×10^{-23}
95% C.I.	$[8.97, 10.0] \times 10^{-8}$	$[1.54, 1.80] \times 10^{-15}$	$[2.49, 3.02] \times 10^{-23}$
Time Taken (s)	14	107	368
Total no. successes	114763	122184	121847
Max no. particles	1470	7302	16050

Table 4. $\lambda = 1, \mu_1 = 2, \mu_2 = 3$, asymptotically optimal scheme.

As expected, these results show a vast improvement over those obtained in [8]. Finally the tables below show the results of numerical simulation for the same problem with a splitting algorithm based on the subsolution $\bar{W}(x) = 0.95W(x)$.

n	10	20	30
Theoretical Value	9.64×10^{-8}	1.60×10^{-15}	2.64×10^{-23}
Estimate	9.57×10^{-8}	1.55×10^{-15}	2.47×10^{-23}
Std. Err.	0.21×10^{-8}	0.05×10^{-15}	0.13×10^{-23}
95% C.I.	$[9.17, 9.98] \times 10^{-8}$	$[1.45, 1.65] \times 10^{-15}$	$[2.22, 2.72] \times 10^{-23}$
Time Taken (s)	8	38	67
Total no. successes	47268	18875	7425
Max no. particles	433	545	703

Table 5. $\lambda = 1, \mu_1 = 3, \mu_2 = 2$, asymptotically suboptimal scheme.

n	10	20	30
Theoretical Value	9.64×10^{-8}	1.60×10^{-15}	2.64×10^{-23}
Estimate	9.32×10^{-8}	1.54×10^{-15}	2.56×10^{-23}
Std. Err.	0.30×10^{-8}	0.07×10^{-15}	0.20×10^{-23}
95% C.I.	$[8.73, 9.92] \times 10^{-8}$	$[1.39, 1.70] \times 10^{-15}$	$[2.16, 2.96] \times 10^{-23}$
Time Taken (s)	7	30	57
Total no. successes	46037	18859	7710
Max no. particles	780	1724	1712

Table 6. $\lambda = 1, \mu_1 = 2, \mu_2 = 3$, asymptotically suboptimal scheme.

The choice of $W(x) = \gamma - \rho_1 x_1 - \rho_2 x_2$ as importance function may seem arbitrary, however it turns out to be a very natural choice. Given $\alpha > 0$ consider a "nice" set B such that for the importance function $W_\alpha(x) = \alpha - \rho_1 x_1 - \rho_2 x_2$, $B \cap \{x : W_\alpha(x) > 0\} = \emptyset$ and $B \cap \{x : W_\alpha(x) = 0\} \neq \emptyset$. Then

$$\lim_{n\to\infty} -\frac{1}{n} \log p^n = \alpha,$$

where

$$p^n = P_{(0,0)}(\text{queue reaches } nB \text{ before first return to } (0,0)).$$

Intuitively this means that all points on a level set of the function $W_\alpha(x)$ have the same asymptotic probability. Thus given any such nice set B we can identify its large deviations rate by finding the unique α^* such that $B \cap \{x : W_{\alpha^*}(x) > 0\} = \emptyset$ and $B \cap \{x : W_{\alpha^*}(x) = 0\} \neq \emptyset$. Further $W_{\alpha^*}(x)$ will be an asymptotically optimal importance function.

That the family of functions W_α has such a property is because the stationary probabilities for a stable tandem Jackson network have the product form $\pi(\{i,j\}) = (1-\rho_1)(1-\rho_2)\rho_1^i\rho_2^j$. Indeed, by using an argument based on the recurrence theorem, we can see that every stable tandem Jackson network has a family of affine subsolutions with the same property. Further this will be true for any N-dimensional queueing network for which the stationary probabilities π have asymptotic product form, by which we mean that there exist $\rho_1, \ldots, \rho_N$ such that for any nice set B

$$\lim_{n\to\infty} -\frac{1}{n}\log \pi(nB) = \inf\{x_1\rho_1 + \cdots + x_N\rho_N : (x_1, \ldots, x_N) \in B\}.$$

6.3 Non-Markovian Process

Since many models are non-Markovian we present an example of splitting for a non-Markovian process. Consider a tandem network whose arrival and service rates are modulated by an underlying process M_t which takes values in the set $\{1,2\}$, such that the times taken for the modulating process to switch states are independent exponential random variables with rate $\gamma(1)$ if M is in state 1 and $\gamma(2)$ otherwise. Let $\lambda(1), \mu_1(1), \mu_2(1)$ and $\lambda(2), \mu_1(2), \mu_2(2)$ be the service rates of the queue in the first and second states respectively. It is known (see, e.g., [4]) that the Hamiltonian can be characterized in terms of the solution to an eigenvalue/eigenvector problem parameterized by p. This characterization is used for calculating the various roots to $\mathbb{H}(p) = 0$ used below.

Consider again the single shared buffer problem. Let $\lambda(1) = 1, \mu_1(1) = 3.5, \mu_2(1) = 2.5, \gamma(1) = 0.2$ and $\lambda(2) = 1, \mu_1(2) = 4.5, \mu_2(2) = 4.5, \gamma(2) = 0.5$. Using a verification argument, one can show that $W(x) = 1.00029(1 - x_1 - x_2)$ is a subsolution with the maximal value $W(0)$. Thus using $W(x)$ leads to an asymptotically optimal splitting scheme. The results of simulations run using this importance function are shown below, where again each estimate was derived using 20,000 runs.

n	30	40	50
Theoretical Value	6.36×10^{-13}	2.88×10^{-17}	1.30×10^{-21}
Estimate	6.05×10^{-13}	2.91×10^{-17}	1.33×10^{-21}
Std. Err.	0.21×10^{-13}	0.11×10^{-17}	0.06×10^{-21}
95% C.I.	$[5.63, 6.47] \times 10^{-13}$	$[2.69, 3.13] \times 10^{-17}$	$[1.21, 1.44] \times 10^{-21}$
Time Taken (s)	3	5	8
Total no. successes	47992	51023	51434
Max no. particles	195	330	469

Table 7. Markov-modulated network, total population overflow.

It is also worth revisiting the separate buffers problem for the same queueing network. For the same arrival and service rates one can again use a verification argument to show that $W(x) = 2.2771 - 1.2953x_1 - 0.9818x_2$ leads to an asymptotically optimal splitting scheme. Results of a simulation using 20,000 runs are shown below.

n	10	20	30
Theoretical Value	8.36×10^{-10}	1.07×10^{-19}	1.39×10^{-29}
Estimate	8.37×10^{-10}	1.07×10^{-19}	1.44×10^{-29}
Std. Err.	0.20×10^{-10}	0.03×10^{-19}	0.05×10^{-29}
95% C.I.	$[7.99, 8.76] \times 10^{-10}$	$[1.01, 1.13] \times 10^{-19}$	$[1.34, 1.54] \times 10^{-29}$
Time Taken (s)	12	89	273
Total no. successes	129643	128335	133658
Max no. particles	1215	5126	8051

Table 8. Markov-modulated network, simultaneous separate buffer overflow.

Finally we investigate what happens in this case if we use a strict subsolution as importance function. The table below shows the results of a simulation using 20,000 runs based on the importance function $\bar{W} = 0.95W$.

n	10	20	30
Theoretical Value	8.36×10^{-10}	1.07×10^{-19}	1.39×10^{-29}
Estimate	8.33×10^{-10}	1.14×10^{-19}	1.38×10^{-29}
Std. Err.	0.21×10^{-10}	0.04×10^{-19}	0.07×10^{-29}
95% C.I.	$[7.91, 8.74] \times 10^{-10}$	$[1.06, 1.22] \times 10^{-19}$	$[1.25, 1.51] \times 10^{-29}$
Time Taken (s)	8	42	91
Total no. successes	77221	49032	27527
Max no. particles	883	2887	3800

Table 9. Markov-modulated network, asymptotically suboptimal scheme.

6.4 Rare Events for the Sample Mean

It is also worth noting that this approach works just as well for finite time problems. Assume that $X_1, X_2, \ldots$ is a sequence of i.i.d. $\mathrm{N}(0, I^N)$ random variables where I^N is the N-dimensional identity matrix and let $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Suppose that we are interested in simulating the sequence of probabilities

$$p^n = P\{S_n \in C\}$$

for some set C such that $\bar{C}$ does not include the origin. For $j \in \{1, \ldots, n\}$ let $S_n(j) = \frac{1}{n}\sum_{i=1}^{j} X_i$. Then given sequences x_n, j_n and $x \in \mathbb{R}^N$, $t \in [0,1]$ such that $\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} j_n/n = t$, the large deviations result

$$\lim_{n\to\infty} -\frac{1}{n}\log P\{S_n \in C | S_n(j_n) = x_n\} = W(x,t)$$

holds. Further the PDE corresponding to solutions of the calculus of variations problem is (see [5])

$$W_t + \inf_{\beta} \mathbb{H}(DW;\beta) = 0,$$

where $\mathbb{H}(s;\beta) = \langle s, \beta\rangle + L(\beta)$ and $L(\beta) = \|\beta\|^2/2$. We can put this into the general framework in the standard way, i.e., by considering the time variable as simply another state variable. The set B, for example, is then $C \times \{1\}$. Strictly speaking this problem does not satisfy the conditions used previously, since the sets A and B no longer have disjoint closure. Although we omit the details, it is not difficult to work around this problem.

It is easy to see that any affine function of the form

$$\bar{W}(x,t) = -\langle \alpha, x\rangle + \|\alpha\|^2 - (1-t)H(\alpha),$$

where $H(\alpha) = \|\alpha\|^2/2$, is a subsolution, though it may not have the optimal value at $(0,0)$ and may not be less than or equal to zero on B. We can use the fact that the minimum of a collection of subsolutions is also a subsolution to build a subsolution which satisfies the boundary condition and has the maximal value at $(0,0)$. For example, suppose that $C = \{x \in \mathbb{R}^2 : \langle p_1, x\rangle \geq 1\} \cup \{x \in \mathbb{R}^2 : \langle p_2, x\rangle \geq 1\}$ where $p_1 = (0.6, 0.8)$ and $p_2 = (0.6, -0.8)$. Let $W_1(x) = 1 - \langle p_1, x\rangle - \frac{1}{2}(1-t)$, $W_2(x) = 1 - \langle p_2, x\rangle - \frac{1}{2}(1-t)$. Then $\bar{W} = W_1 \wedge W_2$ is a subsolution and in fact provides an asymptotically optimal splitting scheme since $\bar{W}(0,0) = W(0,0)$. Numerical results are shown below. Each estimate was derived using 100,000 runs. In contrast to all the previous examples where the process evolves on a grid, the simulated process in this case may cross more than one splitting threshold in a single discrete time step. This appears to increase the variance somewhat (at least if the straightforward implementation as described in Section 2 is used), and hence we increased the number of runs to keep the relative variances comparable.

n	20	30	40
Theoretical Value	7.75×10^{-6}	4.33×10^{-8}	2.54×10^{-10}
Estimate	7.36×10^{-6}	4.21×10^{-8}	2.46×10^{-10}
Std. Err.	0.21×10^{-6}	0.12×10^{-8}	0.08×10^{-10}
95% C.I.	$[6.94, 7.78] \times 10^{-6}$	$[3.97, 4.45] \times 10^{-8}$	$[2.31, 2.61] \times 10^{-10}$
Time Taken (s)	15	38	85
Total no. successes	16209	13756	11934
Max no. particles	155	379	298

Table 10. Sample mean for sums of iid.

References

[1] M. Bardi and I. Capuzzo-Dolcetta. *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*. Birkhäuser, 1997.

[2] P.T. de Boer. Some observations on importance sampling and RESTART. In *Proc. of the 6th International Workshop on Rare Event Simulation*, Bamberg, Germany, 2006.

[3] P. Dupuis, A. Sezer, and H. Wang. Dynamic importance sampling for queueing networks. *Ann. Appl. Probab.*, to appear, 2007.

[4] P. Dupuis and H. Wang. Dynamic Importance Sampling for Uniformly Recurrent Markov Chains. *Ann. Appl. Prob.*, 15:1–38, 2005.

[5] P. Dupuis and H. Wang. Subsolutions of an Isaacs equation and efficient schemes for importance sampling. *Math. Oper. Res.*, to appear, 2007.

[6] W. H. Fleming and H. M. Soner. *Controlled Markov Processes and Viscosity Solutions*. Springer-Verlag, New York, 1992.

[7] M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer-Verlag, New York, 1984.

[8] P. Glasserman, P. Heidelberger, P. Shahabuddin, and T. Zajic. A large deviations perspective on the efficiency of multilevel splitting. *IEEE Trans. Automat. Control*, 43:1666–1679, 1998.

[9] P. Glasserman and S. Kou. Analysis of an importance sampling estimator for tandem queues. *ACM Trans. Modeling Comp. Simulation*, 4:22–42, 1995.

[10] P. Glasserman and Y. Wang. Counter examples in importance sampling for large deviations probabilities. *Ann. Appl. Prob.*, 7:731–746, 1997.